ISBN 82-553-0613-7 1986
9/9-1986 No 3

CONFIDENCE INTERVALS AND CONFIDENCE BANDS FOR THE CUMULATIVE HAZARD RATE FUNCTION AND THEIR SMALL SAMPLE PROPERTIES

by

Ole Bie[1], Ørnulf Borgan[2]
and Knut Liestøl[1]

1) Institute of Informatics, University of Oslo
2) Institute of Mathematics, University of Oslo

Abstract

The construction of pointwise confidence intervals and simultaneous confidence bands for the cumulative hazard rate function, based on right censored survival data, is considered. Intervals and bands based on the Nelson-Aalen estimator and certain transforms of this estimator are derived using asymptotic theory. These intervals and bands are illustrated by data on the survival of salmons suffering from the Hitra disease, and their small sample performance are studied by Monte Carlo simulations. The non-transformed intervals and (particularily) bands tend to achieve too low confidence levels. A great improvement is obtained by applying the transformed intervals and bands.

Keywords: Censored survival data; confidence intervals; confidence bands; hazard rate function; Nelson-Aalen estimator; small sample properties.

1. Introduction

Let X be a positive random variable (lifetime) with absolutely continuous distribution function F, survival function $S = 1-F$ and intensity, or hazard rate function, $\alpha = F'/S$. The cumulative hazard rate function is $A(t) = \int_0^t \alpha(u)du$. Nonparametric estimates of this function is useful, e.g. for assessing the validity of model assumptions in survival analysis. In this paper we develop pointwise confidence intervals and simultaneous confidence bands for A, based on right censored survival data, and study their small-sample properties.

Our confidence intervals and bands are based on the empirical cumulative intensity estimator, or the Nelson-Aalen estimator. This estimator is given as follows. Let $X_1,\ldots,X_n$ be independent, each having the same distribution as X. We consider the set-up with right censoring, where X_i is only observed exactly if it does not exceed a (possibly random) censoring time Z_i. Thus, the observed data are $(\tilde{X}_i, D_i)$; $i=1,2,\ldots,n$; where $\tilde{X}_i = X_i \wedge Z_i$, and $D_i = I\{\tilde{X}_i = X_i\}$. Here $s \wedge t = \min\{s,t\}$ and $I\{\cdot\}$ is the indicator function. Let

$$Y_n(t) = \sum_{i=1}^{n} I\{\tilde{X}_i \geq t\} \qquad (1.1)$$

denote the number at risk at $t-$. Then

$$\hat{A}_n(t) = \sum_{\{i:\tilde{X}_i \leq t, D_i=1\}} [Y_n(\tilde{X}_i)]^{-1} \qquad (1.2)$$

is the Nelson-Aalen estimator.

This estimator was proposed independently by Nelson (1969, 1972) and Altshuler (1970). Later Aalen (1978) generalized the estimator to counting process models, and showed how the theory of counting processes, martingales and stochastic integrals can be used to study its statistical properties. Among other things, Aalen showed that the Nelson-Aalen estimator is almost unbiased, with a variance that may be estimated almost unbiasedly by

$$\hat{\sigma}_n^2(t) = \sum_{\{i:\tilde{X}_i \leq t, D_i=1\}} [Y_n(\tilde{X}_i)]^{-2}. \tag{1.3}$$

For a review of these results, see Andersen and Borgan (1985).

In Section 2 below, we review the large sample properties of the Nelson-Aalen estimator, and show how these may be used to develop various pointwise confidence intervals and simultaneous confidence bands. These intervals and bands are illustrated in Section 3 on data from a controlled trial. The trial assessed the effect of vitamin E and selenium on the survival of salmons suffering from the Hitra disease. Section 4 describes the survival and censoring distributions used in the simulations and gives a brief summary of the applied Monte Carlo techinque. The results and comparisons for the pointwise confidence intervals and the simultaneous confidence bands are presented in the Sections 5 and 6. Some concluding remarks are given in Section 7.

2. Pointwise confidence intervals and simultaneous confidence bands. Asymptotic theory

To be able to derive confidence intervals and bands for the cumulative hazard rate function, we first review some large sample results for the Nelson-Aalen estimator. These results presuppose that the censoring is <u>independent</u> in the sense of Kalbfleisch and Prentice (1980, p. 120), see also Gill (1980, Theorem 3.1.1). This is the case for all the usual types of right censoring, like random censorship and (progressive) censoring of Type I and II (Gill, 1980, Corollary 3.1.1). Moreover, we assume that there exists a constant $T > 0$ and a function y with $y(T) > 0$ such that $Y_n(t)$ defined in (1.1) satisfies

$$\sup_{t\in[0,T]} |Y_n(t)/n - y(t)| \xrightarrow{P} 0 \qquad (2.1)$$

as $n \to \infty$. (Note that this implies that $S(T) > 0$, or equivalently that $A(T) < \infty$.)

Then, by the results reviewed by Andersen and Borgan (1985, Appendix), the following holds true:

(I) <u>The Nelson-Aalen estimator (1.2) is uniformly consistent, i.e.</u>

$$\sup_{t\in[0,T]} |\hat{A}_n(t)-A(t)| \xrightarrow{P} 0, \quad \underline{\text{as}} \quad n \to \infty .$$

(II) <u>Let</u> U <u>be a mean zero Gaussian process with</u> $U(0) = 0$, <u>and</u> $\mathrm{Cov}(U(s),U(t)) = \sigma^2(s\wedge t)$ <u>with</u>

$$\sigma^2(t) = \int_0^t \{\alpha(s)/y(s)\}ds. \qquad (2.2)$$

<u>Then</u>

$$\sqrt{n}(\hat{A}_n-A) \xrightarrow{D} U, \quad \underline{\text{as}} \quad n \to \infty,$$

<u>where the weak convergence takes place in the space</u> $D[0,T]$ (cf. Billingsley, 1968).

(III) <u>With</u> $\hat{\sigma}_n^2(t)$ <u>defined by (1.3) and</u> $\sigma^2(t)$ <u>by (2.2) we have</u>

$$\sup_{t\in[0,T]} |n\hat{\sigma}_n^2(t)-\sigma^2(t)| \xrightarrow{P} 0, \quad \underline{\text{as}} \quad n \to \infty.$$

In the rest of Section 2, (I)-(III) are assumed to hold true, and they will be used repeatedly without explicit reference.

We now turn to the construction of pointwise confidence intervals for $A(t)$ for a fixed $t \in [0,T]$. The standard $100(1-\alpha)$ per cent interval is

$$\hat{A}_n(t) \pm c_{\alpha/2} \hat{\sigma}_n(t), \tag{2.3}$$

where $c_{\alpha/2}$ is the upper $\alpha/2$-fractile of the standard normal distribution. It turns out (Section 5 below) that this interval is unsatisfactory for small sample sizes, however, and we therefore consider transformations to improve the approximation.

This well known idea works as follows. If g is a function which is differentiable in a neighbourhood of $A(t)$, and $g'(x)$ is continuous and different from zero at $x = A(t)$, then by the δ-method

$$\frac{g(\hat{A}_n(t)) - g(A(t))}{|g'(\hat{A}_n(t))|\hat{\sigma}_n(t)} \xrightarrow{D} N(0,1)$$

as $n \to \infty$ (e.g. Serfling, 1980, p. 118-119). It follows that a $100(1-\alpha)$ per cent confidence interval for $g(A(t))$ is

$$g(\hat{A}_n(t)) \pm c_{\alpha/2}|g'(\hat{A}_n(t))|\hat{\sigma}_n(t). \tag{2.4}$$

For a given function g this interval may be converted into a $100(1-\alpha)$ per cent interval for $A(t)$.

In this paper we consider the transformations $g(x) = \log x$ and $g(x) = \arcsin(e^{-x/2})$, the latter being variance stabilizing for the situation with no censoring. These transformations correspond to the log-minus-log transformation and the arc sine-square root transformation suggested by Kalbfleisch and Prentice (1980, pp. 14-15) and Nair (1984), respectively, for the survival function based on the Kaplan-Meier estimator. The logarithmic-transformation gives the $100(1-\alpha)$ per cent confidence interval

$$\hat{A}_n(t)\, e^{\pm c_{\alpha/2}\hat{\sigma}_n(t)/\hat{A}_n(t)} \tag{2.5}$$

while the arc sine-transformation gives the interval

$$- 2\log[\sin\{\min(\pi/2,\ \arcsin(e^{-\hat{A}_n(t)/2}) + \tfrac{1}{2}c_{\alpha/2}\hat{\sigma}_n(t)(e^{\hat{A}_n(t)}-1)^{-\frac{1}{2}})\}]$$

$$\leq A(t) \leq$$

$$\leq - 2\log[\sin\{\max(0,\arcsin(e^{-\hat{A}_n(t)/2}) - \tfrac{1}{2}c_{\alpha/2}\hat{\sigma}_n(t)(e^{\hat{A}_n(t)}-1)^{-\frac{1}{2}})\}]. \tag{2.6}$$

We then consider the construction of simultaneous confidence bands for A on subintervals of $[0,T]$. A class of such bands may be derived as follows (e.g. Doksum and Yandell, 1984). Let q be a continuous and nonnegative function on $[t_1,t_2]$ where $0 \leq t_1 \leq t_2 \leq T$. Then

$$\frac{\sqrt{n}(\hat{A}_n-A)}{1+n\hat{\sigma}^2_n}\, q\circ\frac{n\hat{\sigma}^2_n}{1+n\hat{\sigma}^2_n} \xrightarrow{D} \frac{U}{1+\sigma^2}\, q\circ\frac{\sigma^2}{1+\sigma^2}$$

on $D[t_1,t_2]$ as $n \to \infty$, where U and σ^2 are defined in (II) above, and $\circ$ denotes composition. Let W^0 be the standard Brownian bridge. Then the processes $(U/(1+\sigma^2))q\circ(\sigma^2/(1+\sigma^2))$ and $(qW^0)\circ(\sigma^2/(1+\sigma^2))$ have the same distribution, both being zero mean Gaussian processes with the same covariance function. It follows that

$$\sup_{t\in[t_1,t_2]}\left|\frac{\sqrt{n}(\hat{A}_n(t)-A(t))}{1+n\hat{\sigma}^2_n(t)}\, q\left(\frac{n\hat{\sigma}^2_n(t)}{1+n\hat{\sigma}^2_n(t)}\right)\right| \xrightarrow{D} \sup_{x\in[c_1,c_2]}|q(x)W^0(x)|,$$

as $n \to \infty$, where

$$c_i = \sigma^2(t_i)/(1+\sigma^2(t_i)) \tag{2.7}$$

for $i = 1,2$.

This result may be inverted to yield simultaneous confidence bands. More precisely the $100(1-\alpha)$ per cent confidence band for A on $[t_1,t_2]$ is

$$\hat{A}_n(t) \pm \frac{K_{q,\alpha}(c_1,c_2)}{\sqrt{n}}(1+n\hat{\sigma}_n^2(t))/q\left(\frac{n\hat{\sigma}_n^2(t)}{1+n\hat{\sigma}_n^2(t)}\right) \tag{2.8}$$

with $K_{q,\alpha}(c_1,c_2)$ the upper α fractile in the distribution of $\sup_{x\in[c_1,c_2]}|q(x)W^0(x)|$. For situations where one or both of c_1 and c_2 are not known, the unknown c_i may be replaced by

$$\hat{c}_i = n\hat{\sigma}_n^2(t_i)/(1+n\hat{\sigma}_n^2(t_i)) \tag{2.9}$$

in (2.8).

Two choices of the weight function q in (2.8) seem to be of particular interest. The choice $q_1(x) = \{x(1-x)\}^{-\frac{1}{2}}$ yields confidence bands proportional to the pointwise one (2.3). With this choice we may consider confidence bands on $[t_1,t_2]$, where $0<t_1<t_2<T$ are such that $0<c_1<c_2<1$, cf. (2.7). The resulting $100(1-\alpha)$ per cent confidence band for A on $[t_1,t_2]$ is

$$\hat{A}_n(t) \pm d_\alpha(\hat{c}_1,\hat{c}_2)\hat{\sigma}_n(t), \tag{2.10}$$

where $d_\alpha(c_1,c_2) = K_{q_1,\alpha}(c_1,c_2)$ is the upper α fractile in the distribution of $\sup_{c_1<x<c_2}|W^0(x)/\sqrt{x(1-x)}|$. This fractile may be found by the asymptotic approximation (Miller and Siegmund, 1982, formula (8))

$$P\{\sup_{c_1<x<c_2}\left|\frac{W^0(x)}{\{x(1-x)\}^{\frac{1}{2}}}\right| > d\}$$

$$= \frac{4\phi(d)}{d} + \phi(d)(d-\tfrac{1}{d})\log\{\frac{c_2(1-c_1)}{c_1(1-c_2)}\} + o\{\phi(d)/d\},$$

as $d \to \infty$, where $\phi(d)$ is the standard normal density

$(2\pi)^{-\frac{1}{2}}\exp(-\frac{1}{2}d^2)$. We note that (2.10) is equivalent to the confidence band proposed by Hjort (1985) based on a transformation to the Ornstein-Uhlenbeck process. Following Nair (1984), who considered a band similar to our (2.10) for the survival function, we will denote (2.10) the (nontransformed) equal probability band or <u>EP-band</u> for short.

An alternative choice of weight function is $q_2(x) = 1$. The resulting $100(1-\alpha)$ per cent confidence band for A on $[t_1,t_2]$ is

$$\hat{A}_n(t) \pm \frac{e_\alpha(\hat{c}_1,\hat{c}_2)}{\sqrt{n}}(1+n\hat{\sigma}_n^2(t)) \tag{2.11}$$

where $e_\alpha(c_1,c_2) = K_{q_2,\alpha}(c_1,c_2)$ is the upper α fractile in the distribution of $\sup\limits_{c_1<x<c_2}|W^0(x)|$. For this band we will typically let $[t_1,t_2]$ be the whole of $[0,T]$, in which case tables of $e_\alpha(c_1,c_2) = e_\alpha(0,c_2)$ are given by Koziol and Byar (1975) and Hall and Wellner (1980) for selected values of α and c_2. An expression for the distribution of $\sup\limits_{0<x<c_2}|W^0(x)|$ is given by Hall and Wellner (1980, formula (2.9)). The band (2.11) is similar to the one proposed by Hall and Wellner (1980) for the survival function, and we will denote (2.11) a (nontransformed) Hall-Wellner type band or <u>HW-band</u> for short.

As was the case for the pointwise confidence intervals, it may also for confidence bands be advantageous to consider transformations to improve the approximation of the asymptotic distribution. Let $0<t_1<t_2<T$ and q be as above, and let g be a function such that $g'(x) \neq 0$ and continuous on $[x_1-\varepsilon, x_2+\varepsilon]$ for an $\varepsilon>0$, where $x_i = A(t_i)$. Then $\sqrt{n}(g\circ\hat{A}_n - g\circ A) \xrightarrow{D} (g'\circ A)U$ on $D[t_1,t_2]$, as $n \to \infty$, where U is given in (II) above (cf. the Appendix). Therefore $\sqrt{n}(g\circ\hat{A}_n - g\circ A)/g'\circ\hat{A}_n$ has the same asymptotic distribution as $\sqrt{n}(\hat{A}_n - A)$ on $[t_1,t_2]$. By the argument leading to

(2.8), it follows that a 100(1-α) per cent confidence band for $g \circ A$ on $[t_1, t_2]$ is

$$g(\hat{A}_n(t)) \pm \frac{K_{q,\alpha}(c_1,c_2)}{\sqrt{n}} |g'(\hat{A}_n(t))| (1+n\hat{\sigma}_n^2(t))/q(\frac{n\hat{\sigma}_n^2(t)}{1+n\hat{\sigma}_n^2(t)}).$$

As for the pointwise interval (2.4), this band may, for a given function g, be converted into a 100(1-α) per cent confidence band for A. For the transformations $g(x) = \log x$ and $g(x) = \arcsin(e^{-x/2})$ we get bands corresponding to (2.5) and (2.6), respectively. For the weight function $q_1(x) = \{x(1-x)\}^{-\frac{1}{2}}$, the bands are found by replacing $c_{\alpha}/2$ by $d_{\alpha}(\hat{c}_1, \hat{c}_2)$ in (2.5) and (2.6). These bands, which we will denote the logarithmic- and arc sine-transformed EP-bands, are valid on $[t_1, t_2]$ where $0<t_1<t_2<T$ are such that $0<c_1<c_2<1$, cf. (2.7). For $q_2(x) = 1$, the bands are found by replacing $c_{\alpha/2}\, \hat{\sigma}_n(t)$ by $n^{-\frac{1}{2}} e_{\alpha}(\hat{c}_1, \hat{c}_2)(1+n\hat{\sigma}_n^2(t))$ in (2.5) and (2.6). These bands will be denoted the logarithmic- and arc sine-transformed HW-bands, and they are valid on $[t_1, t_2]$, where $0<t_1<t_2<T$ with $A(t_1) > 0$. In practice one may here use the (slightly) conservative approximation $e_{\alpha}(c_1, c_2) \approx e_{\alpha}(0, c_2)$, when c_1 is close to zero.

3. An example

The data were obtained from a controlled trial evaluating the effect of vitamin E and selenium in the treatment of a fish desease of unknown etiolgy (the Hitra disease) causing mass death of salmons in Norwegian hatcheries (R. Salte, pers.comm.). We will only consider the results from the group given both vitamin E and selenium. This group initially consisted of 50 fish. At day 2, 20 and 42 five fish, randomly drawn among the survivors, were slaughtered for histological and farmacological examinations.

Moreover, five fish were censored by the closing of the study at day 63. This is an example of independent censoring, and the results presented in Section 2 apply.

In Fig 1a, the Nelson-Aalen estimator is shown together with the standard 95% intervals (2.3) and the corresponding 95% EP-band (2.10). The EP-band is given with $\hat{c}_1 = 0.05$ and $\hat{c}_2 = 0.885$ (cf. (2.9)), corresponding to $t_1 = 7$ and $t_2 = 63$. Fig. 1b shows the logarithmic-transformed intervals (2.5) and the corresponding logarithmic-transformed EP-band, while Fig. 1c shows the arc sine-transformed intervals (2.6) and the corresponding arc sine-transformed EP-band. Note that the transformations induce clearly nonsymmetric intervals and bands. Note also that there are easily detectable differences between the two transformed bands. These differences are most marked during the first days of the trial, where the logarithmic-transformed band has boundaries which are shifted upwards compared to those of the arc sine-transformed band. (Although not easily seen from the figure, the difference is also marked at the lower boundary, e.g. at day 7 the values for the intervals are 0.011 and 0.004, respectively.) Fig 1d shows the HW-band (2.11) and the logarithmic- and arc sine-transformed HW-bands. The lower boundary of the (nontransformed) HW-band as well as the upper boundary of the transformed bands, were not monotone. In case of the (nontransformed) HW-band, the highest value of the lower boundary was extended to the end of the trial, and similarily the lowest values of the upper boundary for the transformed bands were extended back to the time of the first death. Note that the HW-bands are broader than the EP-bands at early and late points in time. Moreover, the transformed bands are clearly nonsymmetric, and the logarithmic-transformed bands are shifted upwards compared to the arc sine-transformed band during the early part of the trial.

4. Survival and censoring distributions, simulation technique

In the Monte Carlo simulations, we adopt the random censorship model, where the censoring times $Z_1,\dots,Z_n$ of Section 1 are independent and identically distributed with absolutely continuous distribution function H, and independent of the X_i's. Then, by the Glivenko-Cantelli theorem, (2.1) is valid with $y = (1-F)(1-H)$, so that (I)-(III) of Section 2 hold true provided that $(1-F(T))(1-H(T)) > 0$.

Various combinations of three survival distributions and two censoring distributions have been simulated. In many respects all the combinations gave similar results, and much of the discussion is therefore restricted to a standard situation where both the survival and the censoring follow a standard exponential distribution (with parameter 1). By changing the parameter of one of the exponential distributions different degrees of censoring may be obtained. Note also that, except for the time scale, these exponential models represent all models of the form $1-H = (1-F)^{\theta}$ for some θ (the Koziol-Green models).

The two other survival distributions are of the Weibull type. The first is a Weibull (1.35,2), i.e. it has survival function $e^{-1.35t^2}$, and it represents a situation where the majority of the deaths occurs at intermediate times (mode approximately 0.6). Combined with a standard exponential censoring distribution, the degree of censoring will be approximately 50%. The second Weibull distribution has survival function $e^{\sqrt{2}t^{0.5}}$ corresponding to an expected life length of 1, but with a large variance, i.e. many early deaths combined with a heavy right tail. This last distribution gave results which did not deviate markedly from those obtained with the exponential distributions, and no results are therefore given in the tables below. The second censoring distribution applied, is a uniform distribution on the interval $[0,1.6]$. Combined with the standard exponential survival distribution, the degree of censoring is close to 50%.

In the simulations, it is advantageous to generate directly $\tilde{X}_{(1)}<\ldots<\tilde{X}_{(n)}$, i.e. the ordered $\tilde{X}_i$'s, and the corresponding $D_{(i)}$'s. Let $\alpha = F'/(1-F)$ and $\beta = H'/(1-H)$ be the hazard rate functions corresponding to the survival and censoring distributions. Then, given the i-th censored lifetime $\tilde{X}_{(i)}$ (with $\tilde{X}_{(0)}=0$), the next censored life time $\tilde{X}_{(i+1)}$ is generated as follows. Two pseudo-random variates X'_{i+1} and Z'_{i+1} are generated (by the inversion principle) with hazard rate functions equal to $(n-i)\alpha(t)$ and $(n-i)\beta(t)$, repectively, for $t > \tilde{X}_{(i)}$, and zero otherwise. Then $\tilde{X}_{(i+1)} = X'_{i+1} \wedge Z'_{i+1}$ and $D_{(i+1)} = I\{\tilde{X}_{(i+1)}=X'_{i+1}\}$. It is a simple exercise to see that this generation procedure produces the same results as does the direct generation of $\tilde{X}_i$'s and D_i's, which would then have to be sorted. However, this generation procedure gives a dramatic reduction in computer time. The computer programs were written in the language C, and run on a VAX 11/780.

The simulated results for the confidence intervals and bands with nominal level of confidence 90% and 95% are based on 10 000 replications, corresponding to a standard error on the estimated error rates of (approximately) 0.003 and 0.002, respectively. (The error rate of an interval or band with level of confidence $1-\alpha$ is α.) The simulations for the confidence intervals with nominal level of confidence 99% are based on 20 000 replications, corresponding to a standard error of (approximately) 0.0007 on the estimated error rates.

5. Pointwise confidence intervals. Small sample properties

Table 1 shows the error rates of confidence intervals with nominal confidence level 95% at 5 points in time, for three combinations of survival and censoring distributions and for n equal to 25, 50, 100, and 200. The error rates obtained when applying the standard interval (2.3) are too high, especially for $n = 25$. This is seen at all points in time and for all survival/censoring distributions. (The censoring is in all cases about 50%, see Section 4.) A considerable improvement is obtained by applying one of the transformed intervals. Although the logarithmic-transformed interval (2.5) gives slightly too low error rates, and the arc sine-transformed interval (2.6) slightly too high rates, the achieved confidence levels seem acceptable even for $n = 25$.

A change of the censoring distribution from exponential to uniform gives no essential changes in the results. Similarily, a change of survival distribution from standard exponential to Weibull ($\sqrt{2}$,0.5) gives only small changes in the results (not shown). With the Weibull (1.35,2) survival distribution very high error rates are seen at points early in time. This may be expected, however, since it reflects the fact that almost no deaths have occurred at so low values of t. (For $n = 25$ and $t = 0.2$ the expected number of deaths is only slightly above 1.) At $t = 0.6$, when the expected number of deaths is near 7, the error rates are close to those obtained for the other survival/censoring distributions.

To be completely satisfactory, a confidence interval should fall above and below the true parameter value an approximately equal number of times. Table 2 shows that this is a more difficult task than to achieve an acceptable total error rate. All the intervals produce unbalanced number of failures. However, the transformed intervals are clearly better than the standard one, and the arcsine-transformed interval seems to be somewhat better than the logarithmic-transformed interval. Similar results are obtained at other points in time and with other survival/censoring distributions.

Table 3 shows that the qualitative results found for the 95% confidence intervals carry over to the 90% intervals. However, with a nominal confidence level of 99%, the approximation becomes less satisfactory, implying that a larger number of individuals is needed to achieve acceptable confidence levels.

Table 4 shows results obtained with the standard exponential survival distribution and with an exponential censoring distribution with a parameter value adjusted to obtain various degrees of censoring. When interpreting this table, it should be realized that the Nelson-Aalen estimator is nonparametric and may only attain a finite number of values. This is also the case for the estimator for its variance, cf. (1.3). Because of this discretness it may be impossible to achieve a true level of confidence of exactly 95%. This phenomenon is most pronounced when there is no censoring, since for this situation the Nelson-Aalen estimator and the estimator for its variance may take fewer values than when censoring is present. This explains the unsystematic pattern seen for 0% censoring in Table 4. Neglecting this phenomenon, the table shows the expected decrease in performance with increasing amount of censoring. Note, however, that even with 75% censoring, the transformed intervals produce acceptable results. With the survival and censoring distributions shown in the table, the logarithmic-transformed interval performs better than the arc sine-transformed, but this is less pronounced in other situations.

6. Simultaneous confidence bands. Small sample properties

The pointwise intervals are based on the weak convergence of (transforms of) the Nelson-Aalen estimator for a fixed value of t.

On the other hand, the simultaneous confidence bands are based on the weak convergence of the estimator viewed as a stochastic process. One may expect that a larger number of individuals is needed to ensure that the limiting process is a good approximation, than is the case for a fixed value of t, and that this will be reflected in the performance of the confidence bands. Table 5 shows that this effect is marked for the nontransformed EP-band (2.10) and HW-band (2.11). Both of these bands achieve confidence levels which are far from the nominal ones, even for $n = 200$. Also for the bands, however, the transformations improve the results considerably, and in terms of total error rates the results are acceptable. Note that the results do not depend much on the survival and censoring distributions.

Table 6 illustrates, for $n = 50$, at which points in time the confidence bands fail to cover the true cumulative hazard rate for the first time, and whether they are above or below the true value when they miss. It is seen that the nontransformed bands nearly always are below the true cumulative hazard rate when they miss. For the logarithmic-transformed EP-band a rather high lower boundary (cf. Fig. 1) causes a large number of errors at points early in time. At points later in time the new cases of non-covering bands are caused by a too low upper boundary. The arc sine-transformed EP-band produces the majority of its errors by having a too low upper boundary. The HW-bands concentrate their errors at intermediate points in time (cf. Fig. 1), but otherwise produce results similar to those seen for the EP-bands. These time patterns of the errors are not satisfactory. A restriction of the time interval, over which the bands are computed, improves the results, as does a change to 90% confidence level (shown for the logarithmic-transformed EP-band in Table 7). However, an increase in the number of individuals seems to be neccessary to get completely satisfactory results (Table 7). Note also that in this respect the arc sine-transformed band seems to be somewhat better than the logarithmic-transformed one.

7. Concluding remarks

The most important conclusion of the present study is that confidence intervals and bands for the cumulative hazard rate function, based on right censored survival data, can be markedly improved through the use of transformations. In our opinion, the advantages clearly outweigh the extra calculations needed, and transformed intervals and bands should therefore be used in most cases. We have, however, only considered two transforms, and there may exist better ones.

Considering the confidence intervals, it seems as if about 10 deaths, and a similar number still at risk, are sufficient for the transformed intervals to perform reasonably well, and even with 5 individuals in these categories, the intervals will probably be accurate enough for many purposes. The main difference between the logarithmic- and the arc sine-transformed intervals is that the former tends to be slightly conservative, while the latter tends to achieve levels of confidence slightly below the nominal ones. Corresponding to this, the arc sine-transformed interval tends to be more narrow than the logarithmic-transformed one. Moreover, the logarithmic-transformed interval more often falls above than below the true parameter value, while the opposite is the case for the arc sine-transformed interval. The choice between these two transformed intervals will thus depend on the purpose, and we can see no reason for suggesting any standard choice.

Considering the confidence bands, it is more difficult to state necessary sample sizes. For the transformed bands, the achieved levels of confidence are acceptable with fairly few observations, but the way the bands distribute the errors (above or below the true value, time pattern) may appear less satisfactory. Although there is no unique best way for a band to distribute the errors over time, the observed dependence on the number of observations is undesirable (Table 7). Thus, if only a rough estimate of the uncertainty of the estimated hazard is wanted, a band based on 20-30 deaths (or ever fewer) may be satisfactory, while larger numbers may be necessary for other purposes. It appears as if the arc sine-transformed EP-band may have some advantages over the log--transformed one regarding the distribution of the errors over time. For the EP-bands, better statistical properties can be achieved by restricting the time interval covered by the band. The main difference between the EP-bands and the HW-bands is how the errors are distributed over time, and the choice between them will depend on the time interval over which one wants a narrow band.

Our study has been restricted to the set-up with right censored survival data. However, the Nelson-Aalen estimator may also be used for other purposes, e.g. to estimate the cumulative transition intensities in a Markov chain model (cf. Andersen and Borgan, 1985). With minor modifications the results of Section 2 are valid for these other situations. The small sample properties of the confidence intervals and bands may be different, however, and further studies are needed here.

Acknowledgements

We are grateful to Nils Lid Hjort for clarifying comments and discussions concerning the results on weak convergence of processes in Section 2 and the Appendix. He also made us aware of the important approximation formula of Miller and Siegmund (1982). Some of Ørnulf Borgan's work on this project was done during a visit (June to August 1985) at the Statistical Research Unit in Copenhagen, supported by Johan and Mimi Wessmann's legacy.

Appendix. Weak convergence of $\sqrt{n}\,(go\hat{A}_n - goA)$

We consider the situation of Section 2. Let $0<t_1<t_2<T$, and let g be a function such that $g'(x) \neq 0$ and continuous on $[x_1-\varepsilon, x_2+\varepsilon]$ for an $\varepsilon > 0$, where $x_i = A(t_i)$. We will prove that

$$\sqrt{n}\,(go\hat{A}_n - goA) \xrightarrow{D} (g'oA)U \tag{A.1}$$

on $D[t_1,t_2]$, as $n \to \infty$, where U is the Gaussian process defined in (II) of Section 2. Our proof is extracted from the proof of Theorem A.1 in Borgan and Gill (1982).

To prove (A.1) we apply a Skorohod construction (cf. Breslow and Crowley, 1974) to the sequence $\{\sqrt{n}(\hat{A}_n-A)\}$ of random elements of $D[t_1,t_2]$ which converges weakly to U. Thus we may replace $\{\sqrt{n}(\hat{A}_n-A)\}$ and U by random functions, defined on a new sample space, having the same distribution for each n, but which also satisfy $\sqrt{n}\,(\hat{A}_n-A) \to U$ almost surely, as $n \to \infty$, in the Skorohod topology. (We use the same notation for these new processes.) Since U is continuous, we actually have almost sure convergence in the supremum norm $\|\cdot\|$ on $D[t_1,t_2]$. On this new sample space we shall prove

$$\|\sqrt{n}(go\hat{A}_n - goA)-(g'oA)U\| \to 0 \tag{A.2}$$

almost surely, as $n \to \infty$, which will imply (A.1) in the original set-up.

Now g' is bounded and uniformly continuous on $[x_1-\varepsilon, x_2+\varepsilon]$. By the mean value theorem we have on the event

$$\{\hat{A}_n(t) \in [x_1-\varepsilon, x_2+\varepsilon]\}$$

that

$$\sqrt{n}\,(g(\hat{A}_n(t))-g(A(t))) = g'(\xi_n(t))\,\sqrt{n}\,(\hat{A}_n(t)-A(t)),$$

where $\xi_n(t)$ is between $\hat{A}_n(t)$ and $A(t)$. Therefore, on this event

$$\|\sqrt{n}(g\circ\hat{A}_n - g\circ A)-(g'\circ A)U\|$$

$$\leq \|g'\circ\xi_n - g'\circ A\|\cdot\|\sqrt{n}(\hat{A}_n - A)\|$$

$$+ \|g'\circ A\|\cdot\|\sqrt{n}(\hat{A}_n - A)-U\| .$$

Now, $A_n \in [x_1-\varepsilon, x_2+\varepsilon]$ almost surely for all large enough n, which proves (A.2).

References

Aalen, O.O. (1978). Non-parametric inference for a family of counting processes. Ann. Statist. 6, 701-726.

Altshuler, B. (1970). Theory for the measurement of competing risks in animal experiments. Math. Biosc. 6, 1-11.

Andersen, P.K. & Borgan, Ø. (1985). Counting process models for life history data: A review (with discussion). Scand. J. Statist. 12, 97-158.

Billingsley, P. (1968). Convergence of probability measures. Wiley, New York.

Borgan, Ø & Gill, R.D. (1982). Case-control studies in a Markov chain setting. Preprint SW 89/82. Math. Centre, Amsterdam.

Breslow, N.E. & Crowley, J. (1974). A large sample study of the life table and product limit estimates under random censorship. Ann. Statist. 2, 437-453.

Doksum, K.A. & Yandell, B.S. (1984). Tests for exponentiality. Pp. 579-611 in Handbook of Statistics, Vol. 4 (eds. P.R. Krishnaiah and P.K. Sen). Elsevier Science Publishers.

Gill, R.D. (1980). Censoring and stochastic integrals. Math. Centre Tracts 124, Math. Centre, Amsterdam.

Hall, W.J. & Wellner, J.A. (1980). Confidence bands for a survival curve with censored data. Biometrika 67, 133-143.

Hjort, N.L. (1985). Contribution to the discussion of a paper by Andersen & Borgan. Scand. J. Statist. 12, 97-158.

Kalbfleisch, J.D. & Prentice, R.L. (1980). The statistical analysis of failure time data. Wiley, New York.

Koziol, J.A. & Byar, D.P. (1975). Percentage points of the asymptotic distribution of one and two sample K-S statistics for truncated or censored data. Technometrics 17, 507-510.

Miller, R. & Siegmund, D. (1982). Maximally selected chi-square statistics. Biometrics 38, 1011-1016.

Nair, V.N. (1984). Confidence bands for survival functions with censored data: A comparative study. Technometrics 26 265-275.

Nelson, W. (1969). Hazard plotting for incomplete failure data J.Qual. Technol. 1, 27-52.

Nelson, W. (1972). Theory and application of hazard plotting for censored failure data. Technometrics 14, 945-966.

Serfling, R.J. (1980). Approximation theorems of mathematical statistics. Wiley, New York.

Legend to Figure 1:

Figure 1. Confidence intervals and confidence bands for the cumulative hazard function applied to a set of data concerning the survival of salmons affected by the Hitra disease (see text). In all subfigures the fully drawn line represents the Nelson-Aalen estimator. In 1a the inner pair of lines (•••) represents the non-transformed pointwise confidence limits, and the outer pair (---) the nontransformed EP-band. In 1b the inner pair is the logarithmic-transformed confidence intervals, and the outer pair is the logarithmic-transformed EP-band. 1c shows the arc sine-transformed intervals (inner pair) and the arc sine-transformed EP-band (outer pair). In 1d the three different HW-bands are compared: the non-transformed band (-•-•-), the logarithmic-transformed one (••••••) and the arc sine-transformed one (----).

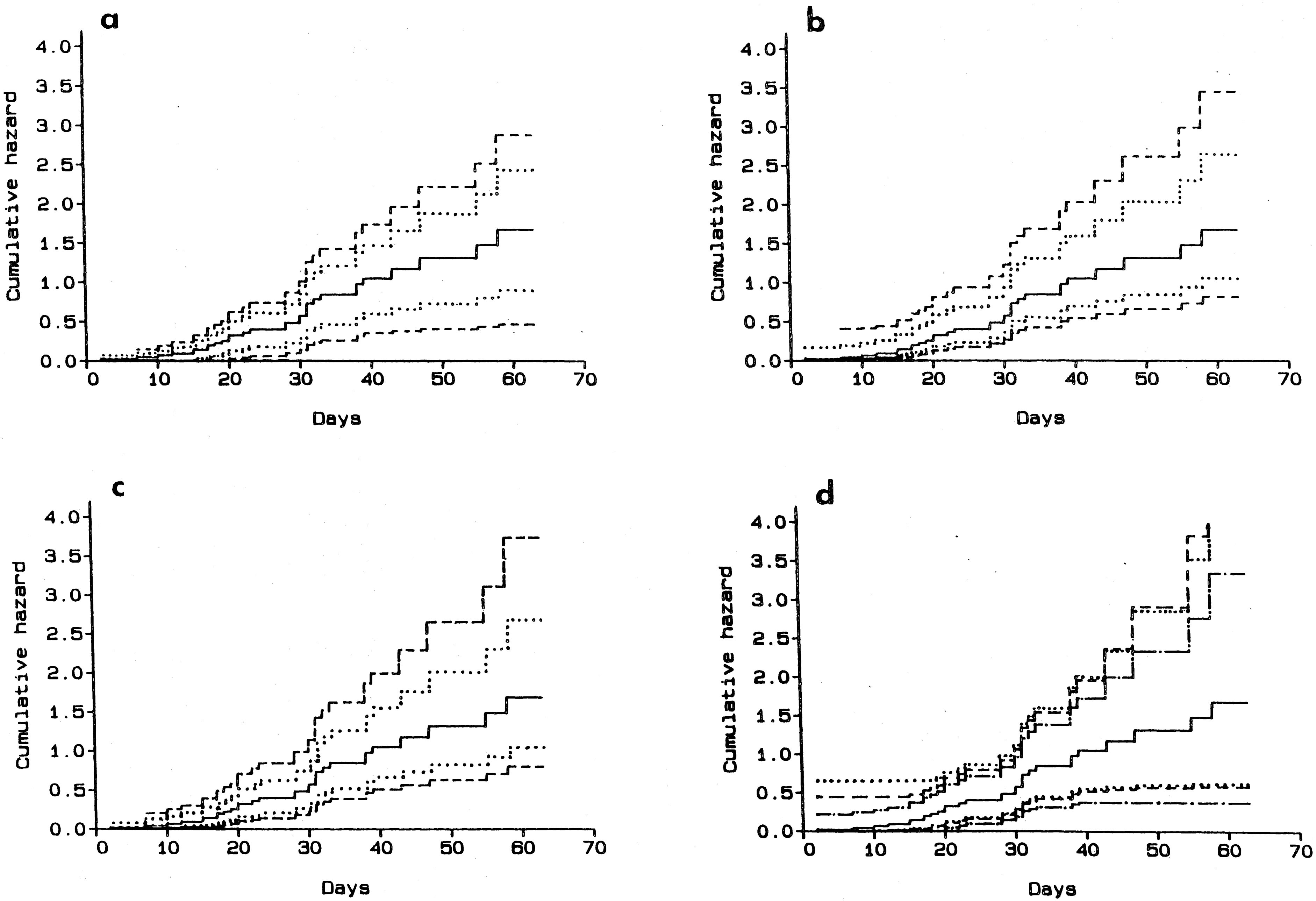
a
b
c
d
Cumulative hazard
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
Days
0
10
20
30
40
50
60
70

Table 1. Achieved error rates of confidence intervals with nominal level of confidence 95%

		Exponential/Exponential[a]					Exponential/Uniform[b]					Weibull/Exponential[c]				
	t =	0.2	0.4	0.6	0.8	1.0	0.2	0.4	0.6	0.8	1.0	0.2	0.4	0.6	0.8	1.0
Standard	n= 25	.08	.08	.07	.07	.09	.08	.08	.07	.07	.08	.31	.10	.08	.08	.09
interval	n= 50	.07	.06	.06	.06	.06	.07	.06	.06	.06	.06	.10	.08	.06	.06	.06
(2.3)	n=100	.06	.06	.06	.06	.06	.07	.06	.06	.05	.06	.09	.07	.06	06	.05
	n=200	.06	.05	.05	.05	.05	.05	.05	.05	.05	.05	.07	.06	.06	.05	.05
Logarithmic-	n= 25	.04	.04	.04	.04	.04	.04	.04	.04	.04	.03	.34	.05	.04	.04	.03
transformed	n= 50	.04	.05	.05	.04	.04	.04	.04	.05	.05	.04	.13	.04	.04	.04	.04
interval	n=100	.05	.05	.05	.05	.05	.05	.05	.05	.05	.05	.05	.05	.05	.05	.05
(2.5)	n=200	.05	.05	.05	.05	.05	.05	.05	.05	.05	.05	.04	.05	.05	.05	.05
Arc sine-	n= 25	.08	.06	.06	.06	.06	.07	.06	.06	.06	.05	.32	.08	.06	.06	.05
transformed	n= 50	.06	.05	.06	.05	.05	.06	.05	.06	.06	.06	.11	.06	.05	.06	.06
interval	n=100	.05	.05	.06	.05	.06	.05	.05	.05	.05	.05	.07	.06	.05	.05	.05
(2.6)	n=200	.05	.05	.05	.05	.05	.05	.05	.05	.05	.05	.06	.05	.05	.05	.05

a) Both survival and censoring distributions are standard exponential.

b) Survival distribution is standard exponential, censoring distribution is uniform over [0,1.6].

c) Survival distribution is Weibull (1.35,2), censoring distribution is standard exponential.

Table 2. The distribution of the errors above and below the true value of the cumulative hazard function at t=0.4 for confidence intervals with nominal level 95%.

		Exponential/ Exponential[a]		Exponential/ Uniform[b]		Weibull/ Exponential[c]	
		Below	Above	Below	Above	Below	Above
Standard interval (2.3)	n= 25	.081	.002	.074	.002	.094	.002
	n= 50	.056	.008	.056	.006	.075	.004
	n=100	.050	.010	.045	.008	.058	.008
	n=200	.038	.015	.038	.015	.045	.012
Logarithmic-transformed interval (2.5)	n= 25	.010	.029	.009	.028	.008	.031
	n= 50	.017	.031	.017	.026	.018	.032
	n=100	.021	.027	.019	.026	.017	.033
	n=200	.020	.028	.020	.029	.020	.028
Arc sine-transformed interval (2.6)	n= 25	.042	.019	.039	.019	.041	.035
	n= 50	.032	.022	.034	.018	.040	.018
	n=100	.033	.022	.030	.020	.036	.021
	n=200	.027	.024	.027	.022	.032	.021

a,b,c) See footnotes of Table 1.

Table 3. Achieved error rates for varying nominal level of confidence. (Survival and censoring are both standard exponential and t=0.4.)

		Nominal level of confidence		
		90%	95%	99%[a]
Standard interval (2.3)	n= 25	.12	.08	.042
	n= 50	.11	.06	.024
	n=100	.10	.06	.017
	n=200	.10	.05	.013
Logarithmic-transformed interval (2.5)	n= 25	.09	.04	.006
	n= 50	.10	.05	.009
	n=100	.10	.05	.009
	n=200	.10	.05	.009
Arc sine-transformed interval (2.6)	n= 25	.12	.06	.016
	n= 50	.11	.05	.014
	n=100	.10	.05	.010
	n=200	.10	.05	.010

a) Based on 20 000 simulations.

Table 4. Achieved error rates with varying degree of censoring for confidence intervals with nominal level of confidence 95%. (Survival and censoring are both standard exponential and t=0.4.)

		Amount of censoring			
		0%	25%	50%	75%
Standard interval (2.3)	n= 25	.05	.07	.08	.11
	n= 50	.07	.06	.07	.08
	n=100	.07	.05	.06	.06
	n=200	.05	.05	.05	.06
Logarithmic-transformed interval (2.5)	n= 25	.05	.04	.04	.05
	n= 50	.05	.05	.05	.05
	n=100	.04	.05	.05	.04
	n=200	.05	.05	.05	.05
Arc sine-transformed interval (2.6)	n= 25	.07	.06	.06	.08
	n= 50	.05	.05	.06	.06
	n=100	.06	.05	.06	.06
	n=200	.05	.05	.05	.05

Table 5. Achieved error rates of confidence bands with nominal level of confidence 95%.

		Exponential/ Exponential[a]	Exponential/ Uniform[b]	Weibull/ Exponential[c]
EP-band (2.10)[d]	n= 25	.19	.16	.21
	n= 50	.17	.12	.20
	n=100	.11	.08	.13
	n=200	.08	.07	.09
Logarithmic-transformed EP-band[d]	n= 25	.06	.05	.06
	n= 50	.06	.05	.06
	n=100	.06	.05	.06
	n=200	.05	.05	.05
Arc sine-transformed EP-band[d]	n= 25	.05	.04	.05
	n= 50	.05	.04	.05
	n=100	.05	.05	.05
	n=200	.05	.05	.05
HW-band (2.11)	n= 25	.17	.13	.20
	n= 50	.15	.11	.16
	n=100	.11	.09	.11
	n=200	.08	.08	.08
Logarithmic-transformed HW-band	n= 25	.06	.05	.06
	n= 50	.06	.05	.06
	n=100	.05	.05	.05
	n=200	.05	.05	.05
Arc sine-transformed HW-band	n= 25	.06	.05	.06
	n= 50	.06	.05	.06
	n=100	.06	.05	.05
	n=200	.05	.05	.05

a, b, c) See footnotes of Table 1

d) The bands are evaluated with $\hat{c}_1=0.05$ and $\hat{c}_2=0.95$.

Table 6. The points in time when confidence bands with nominal level 95% miss the true cumulative hazard function for the first time. (Survival and censoring are both standard exponential and n=50.)

		Time interval[a]						
		0-.2	.2-.4	.4-.6	.6-.8	.8-1.0	1.0-1.2	1.2-
EP-band[b] (2.10)	Below	.016	.040	.019	.012	.009	.010	.041
	Above	.000	.000	.000	.000	.000	.000	.000
Logarithmic-transformed EP-band[b]	Below	.000	.000	.000	.001	.001	.001	.008
	Above	.037	.005	.003	.001	.000	.000	.000
Arc sine-transformed EP-band[b]	Below	.000	.008	.004	.004	.003	.003	.017
	Above	.003	.002	.003	.001	.000	.000	.000
HW-band (2.11)	Below	.000	.010	.024	.027	.021	.019	.058
	Above	.000	.000	.000	.000	.000	.000	.000
Logarithmic-transformed HW-band	Below	.000	.000	.000	.001	.004	.004	.018
	Above	.016	.014	.004	.000	.000	.000	.000
Arc sine-transformed HW-band	Below	.000	.000	.003	.009	.010	.008	.025
	Above	.003	.006	.002	.002	.000	.000	.000

a) Right endpoint not included

b) The bands are evaluated with $\hat{c}_1=0.05$ and $\hat{c}_2=0.95$

Table 7. The effect of varying the nominal level of confidence, the sample size and the time interval on the distribution in time of the first errors of the transformed EP-bands.

		Time interval[a]						
		0-.2	.2-.4	.4-.6	.6-.8	.8-1.0	1.0-1.2	1.2-
Logarithmic-transformed, n= 50, 95%, $\hat{c}_1$=.05, $\hat{c}_2$=.95	Below	.000	.000	.000	.001	.001	.001	.008
	Above	.037	.005	.003	.001	.000	.000	.000
Logarithmic-transformed, n= 50, 95%, $\hat{c}_1$=.10, $\hat{c}_2$=.90	Below	.000	.000	.000	.001	.001	.002	.009
	Above	.027	.005	.002	.002	.000	.000	.000
Logarithmic-transformed, n= 50, 90%, $\hat{c}_1$=.05, $\hat{c}_2$=.95	Below	.000	.000	.001	.001	.003	.002	.014
	Above	.055	.008	.005	.003	.001	.000	.000
Logarithmic-transformed, n= 50, 90%, $\hat{c}_1$=.10, $\hat{c}_2$=.90	Below	.000	.000	.003	.003	.004	.003	.016
	Above	.041	.011	.007	.003	.001	.000	.000
Logarithmic-transformed, n=200, 95%, $\hat{c}_1$=.05, $\hat{c}_2$=.95	Below	.001	.002	.001	.002	.002	.003	.009
	Above	.022	.004	.003	.002	.002	.001	.000
Arc sine-transformed, n=200, 95%, $\hat{c}_1$=.05, $\hat{c}_2$=.95	Below	.007	.005	.003	.003	.003	.003	.011
	Above	.005	.002	.003	.001	.001	.001	.000

a) Right endpoint not included